Exponential Upper Bounds via Martingales for Multiplexers with Markovian Arrivals.

E. Buffet [1, 2] and N.G. Duffield [1, 2].

Abstract. We obtain explicit upper bounds in closed form for the queue length in a slotted time FCFS queue in which the service requirement is a sum of independent Markov processes on the state space $\{0, 1\}$, with integral service rate. The bound is of the form $\mathbb{P}[\text{queue length} \geq b] \leq cy^{-b}$ for any $b \geq 1$ where $c < 1$ and $y > 1$ are given explicitly in terms of the parameters of the model. The model can be viewed as an approximation for the burst-level component of the queue in an ATM multiplexer. We obtain heavy traffic bounds for the mean queue length and show that for typical parameters this far exceeds the mean queue length for independent arrivals at the same load. We compare our results on the mean queue length with an analytic expression for the case of unit service rate, and compare our results on the full distribution with computer simulations.

Key words: Queueing Theory, ATM Multiplexers, Martingales, Upper bounds, Heavy Traffic Limit.

AMS 1991 Classifications: Primary 60K25; Secondary 68M20, 90B22, 68M20

[1] *School of Theoretical Physics, Dublin Institute for Advanced Studies, 10 Burlington Road, Dublin 4, Irleand.*

[2] *School of Mathematical Sciences, Dublin City University, Dublin 9, Ireland.*

1. Introduction.

The queueing theoretic analysis of models of asynchronous transfer mode (ATM) voice multiplexers has been resistant to exact treatment. This is primarily due to the nature of the traffic: a superposition of bursty periodic sources which gives rise to an arrival process which is highly correlated between different times. Such models have attracted much attention recently, predominantly from the point of view of simulation and approximation by simpler models. Existing mathematical treatments of particular models have provided results which are either complicated from the computational point of view (e.g. **[1,11]**) or confined to the asymptotic case of large queue lengths e.g. **[5]**. In this we use a martingale technique to obtain an exponential upper bounds $\mathbb{P}[\text{queue} \geq b] \leq cy^{-b}$ *in closed form* for the queue length in a slotted time model in which the arrival process is a superposition of Markov processes on the state space $\{0,1\}$. Here a line in the state 1 delivers a packet of unit length, while in the state 0 no packet is delivered. Equipped with this bound we obtain an upper bound for the mean queue length, and investigate its asymptotics as the load is increased to the threshold of instability. The bounds show very close agreement with simulation results in large superpositions. Thus the utility of our results is two-fold. Firstly we provide a simple, rigorous upper bound for a multiplexer model, and secondly we give a technique which holds the prospect of generalization to other models.

In a classic note, Kingman **[9]** used techniques from martingale theory in order to obtain exponential bounds for the queue lengths in the queue $GI/G/1$. We briefly present (a slightly modified version of) Kingman's method. Let messages labelled $1, 2, \ldots$ arrive at the queue, and be serviced in a first-come first-served discipline. Set $u_n = r_n - t_n$, where r_n is the service time of the n^{th} message, and t_n is the time between the arrival of the $(n+1)^{\text{th}}$ and n^{th} messages. If the queue is initially empty, then the waiting time of the $(n+1)^{\text{th}}$ message is

$$w_{n+1} = \max\{0, u_n, u_n + u_{n-1}, \ldots, u_1 + u_2 + \ldots u_n\} \quad .$$

Since the u_j are independent and identically distributed, then for any $y \geq 1$ and $b \geq 0$ the event $\{w_n \geq b\}$ has the same probability as the event $\{\max_{1 \leq j \leq n} Y_j \geq y^b\}$ where $Y_j = y^{u_1 + \cdots u_j}$, *provided that* we can choose y such that $\mathbb{E}[y^{u_1}] = 1$. Now since the u_j are i.i.d., then the conditional expectation of Y_{j+1} given the past $u_j, \ldots, u_1$ is

$$\mathbb{E}[Y_{j+1} \mid u_j, \ldots, u_1] = Y_j \quad .$$

In other words, $(Y_j)_{j\geq 1}$ is a martingale. Thus, by the maximal inequality for positive (super)-martingales [**14**], we have the upper bound

$$\mathbb{P}[w_n \geq b] = \mathbb{P}[\max_{1\leq j\leq n} Y_j \geq y^b] \leq y^{-b} \quad .$$

The bound is optimized by taking as $y = \overline{y} := \max_{y\geq 1}\{y \mid \mathbb{E}[y^{u_1}] = 1\}$. That such a $\overline{y} > 1$ exists follows from the fact that $(d/dy)\mathbb{E}[y^{u_1}]_{|y=1} = \mathbb{E}[u_1] < 0$ if the queue is to be stable , while $\mathbb{E}[y^{u_1}] > 1$ for sufficiently large y. The bound is independent of n and so is also a bound for the distribution of $w = \lim_{n\to\infty} w_n$.

Now it is often the case that martingale methods can be used to generalize results which hold for independent random variables to the case of dependent random variables. Indeed, martingale theory has been used in such a way to obtain upper bounds in risk theory (see e.g. [**6**]), although the emphasis here has been finding the analogue of the exponential decay rate y (the safety loading), rather than the best prefactor c. By finding an appropriate exponential martingale of form similar to (Y_j) we are able to find an upper bound of the form $\mathbb{P}[w_n \geq b] \geq cy^{-b}$ for a family of constants $c < 1 < y$ in the case that the u_n are superposition of two-state Markov chains.

More specifically, in Section 2 we give the extension of these methods to obtain a family of upper bounds for the tail of the queue length distribution for the following model. We work in slotted (discrete) time. The service requirement r_n at time each integral time n is equal to the sum of L independent random variables taking values in $\{0, 1\}$, each of which is the state of an independent Markov chain. The rate of service of the queue is $s \in \mathbb{Z}^+$. Such a model can be viewed as approximating the burst component of the queue of bursty periodic sources of period s. Thus r_n represents the arrivals at the queue during a block of s ticks of a multiplexer clock. The imposition of an upper bound for the burst component has been required by other authors in order to bound the cell component of queues in ATM multiplexers [**15**]. Indeed, the present model cannot in any sense be expected to yield predictions about the *cell* component of the multiplexer queue, as investigated in [**16,7**], since the details of the arrival process at the tick level are subsumed within the total arrivals of the block of s ticks. However, comparative simulations amongst various models indicate that the tail distribution of the queue length is reasonably insensitive to the details of the arrival processes, provided that the load and correlations are held constant [**3**], so the simplicity of the model need not be a major problem. Approximation of the present model by a Markov modulated model has been investigated through simulation in [**2**].

We show in Section 3 that amongst the family of bounds, one can be written as $\mathbb{P}[\text{queue } \geq b] \leq cy^{-b}$ for any $b \geq 1$ where $c < 1$ and $y > 1$ are now particular constants which are given explicitly in terms of the parameters of the model. (See equation 3.1). Numerically it appears that this explicit bound is extremely close to the optimal one within the family of bounds when L is large, although some divergence is seen for small L. The prefactor in the explicit bound turns out to be exactly the large deviation bound calculated by Hui [8] in his bufferless resource model. This is the case $b = 0$, where one finds the stationary probability that in any one time slot the number of arriving cells exceeds the service rate s. Thus our bound can be seen as a simple but rigorous interpolation between $b = 0$ and the asymptotic regime $b \to \infty$. We close this section by deriving a bound for the mean queue length, and investigating its behaviour as the traffic load is increased to the threshold of instability.

In Section 4 we compare our bounds in two directions. Firstly, using matrix methods, Viterbi has calculated the mean waiting time for our model in the special case $s = 1$ [17]. For typical parameters of systems which she considered, we find that our upper bound overestimates the waiting time by roughly a factor 2. For larger systems this factor becomes smaller. Secondly, for parameters corresponding to a (scaled down) multiplexer ($s = 40$) we compare our bounds with computer simulations of the same process. Our explicit result is shown to lead to an overestimate of the tail probabilities of the queue, typically by less than an order of magnitude. It seems that the decay constant y is nevertheless well-approximated by the bound.

2. Exponential upper bounds.

We specify our model precisely. There are L independent sources, each of which is represented by a copy of a Markov chain $(X_n)_{n \in \mathbb{Z}}$ taking values in the state space $\{0, 1\}$. Here 0 denotes the silent state, while 1 denotes the active state. The probability of transition from the silent state to the active state is a and from the active to the silent state d. Hence the forward transition matrix for each line is

$$T = \begin{pmatrix} 1-a & a \\ d & 1-d \end{pmatrix}$$

with stationary state $\mu = (a+d)^{-1}(d, a)$. One verifies that the Markov chain is reversible: $\mu_i T_{ij} = \mu_j T_{ji}$ for all $i, j \in \{0, 1\}$.

In terms of lifetimes, on each source the active period is geometrically distributed with parameter $(1-d)$ and so its mean active period is $\sum_{n=1}^{\infty} dn(1-d)^{n-1} = 1/d$ units of time. Likewise the mean silence length on each line is $1/a$ units.

The queue operates as follows. Let $X^\ell, \ell = 1, 2, \ldots, L$ denote the Markov chains of all the input lines. Then at each integral time n, all active lines empty one cell into the buffer of the queue. The queue then services at most s cells from the buffer, s being a fixed quantity. Thus one can write a Lindley equation to relate the queue lengths $(q_n)_{n\in\mathbb{Z}}$ as

$$q_{n+1} = \max\{0, q_n + \tilde{z}_n - s\} \quad ,$$

where $\tilde{z}_n = \sum_{1\le\ell\le L} X_n^\ell$ is the number of sources active at time n. Note that s is a fixed deterministic quantity.

Now $(\tilde{z}_n)_{n\in\mathbb{Z}}$ is clearly itself a stationary ergodic Markov process, and we can write down the elements of its transition matrix:

$$\mathbb{P}[\tilde{z}_{n+1} = p \mid \tilde{z}_n = q] = \sum_{r=0}^{q}\sum_{r'=0}^{L-q} \delta_{p,r+r'} \binom{q}{r}(1-d)^r d^{q-r} \binom{L-q}{r'} a^{r'}(1-a)^{L-q-r'} \quad . \tag{2.1}$$

where $\delta_{p,q} = 1$ if $p = q$ and 0 otherwise. Since the individual line process are reversible, so is $(\tilde{z}_n)$.

Proposition 1. *The waiting time for the model has a unique stationary distribution provided that the following stability condition is satisfied:*

$$a(a+d)^{-1} < \sigma := s/L \quad .$$

Proof: Clearly the workload process $(\tilde{z}_t - s)$ is stationary and ergodic. According to Theorem 3 of the general treatment of Loynes[**12**], any queue with first-come first-served (FCFS) discipline and such a workload process has a unique stationary distribution provided that the mean service demand is less than the mean interarrival time. Since the probability that each line is active at a given moment is $a/(a+d)$ we require then that $La/(a+d) < s$.

■

Thus $L_{max} = s(a+d)/a$ is the maximum number of lines which can be accommodated, and $\rho = L/L_{max} = a/((a+d)\sigma) \le 1$ measures the multiplexer load. We use only $\sigma < 1$, since otherwise $s \ge L$ and there is nothing to prove since the service

rate is never less than the number of arriving packets, and so the (stationary) queue is always empty.

In the following we will calculate the virtual waiting time for this queue. Since the individual packets are deterministic unit length, then the virtual waiting time is identical with the queue length as measured at the end of a time unit. (Thus we do not consider the distribution of waiting times for individual packets within a block arriving during one unit of time). We find it convenient to count time backwards, and use the time reversed Markov chain $z_n = \tilde{z}_{-n}$, i.e. $z_n = \sum_{1\le\ell\le L} X^{\ell}_{-n}$. Then the virtual waiting time at time 0 is

$$Q = \sup_{n\ge 1} Q_n$$

where

$$Q_0 = 0 \qquad \text{and} \qquad Q_n = \max_{1\le m\le n}\{z_1 + \cdots + z_m - ms\} \qquad \text{for } n \ge 1.$$

Define the real functions

$$y(x) := \frac{x\big((1-a)+ax\big)}{(1-d)x+d} \qquad f(x) := \frac{ax+d}{(a+d)x^{\sigma}} \qquad g(x) := \frac{y(x)^{\sigma}}{ax+1-a},$$

and the subsidiary quantity $\alpha(x) := (ax+1-a)^L = y(x)^s/g(x)^L$. For fixed positive numbers x and y we define the sequence of random functions $Y = (Y_k)_{k\in\mathbb{N}}$ by

$$\begin{aligned} Y_1 &= x^{z_1} \\ Y_k &= \alpha(x)^{1-k} y(x)^{z_1+\cdots+z_{k-1}} x^{z_k} \quad \text{for } k > 1 \quad , \end{aligned}$$

where each k let $\mathcal{M}_k$ be the σ-algebra generated by the random variables $\{z_1, \ldots, z_k\}$.

Proposition 2. *$(Y_k)_{k\in\mathbb{N}}$ is a martingale with respect to the filtration $(\mathcal{M}_k)_{k\in\mathbb{N}}$.*

Proof: Due to the reversibility of each of the Markov chains X^{ℓ}, the transition matrix specified in (2.1) is the transition matrix for z_t. Hence

$$\begin{aligned} \mathbb{E}[x^{z_j+1} \mid z_j] &= \sum_{p}^{L}\sum_{r=0}^{z_j}\sum_{r'=0}^{L-z_j} x^p \delta_{p,r+r'} \binom{z_j}{r}(1-d)^r d^{z_j-r}\binom{L-z_j}{r'} a^{r'}(1-a)^{L-z_j-r'} \\ &= \big((1-d)x+d\big)^{z_j}(ax+1-a)^{L-z_j} = \alpha(x)\big(x/y(x)\big)^{z_j} \quad . \end{aligned}$$

Hence for $k \ge 1$

$$\mathbb{E}[Y_{k+1} \mid \mathcal{M}_k] = \alpha(x)^{1-k} y(x)^{z_1+\cdots+z_k}\big(x/y(x)\big)^{z_k} = Y_k \quad .$$

■

We need the following technical result concerning the relative magnitudes of x and $y(x)$.

Proposition 3.

$$y(x) \geq 1 \text{ if } x \geq 1 \quad ,$$

in which case

$$y(x) \leq x \text{ iff } a + d \leq 1 \quad .$$

Proof: First observe from analysis of the quadratic inequality in x obtained by setting $y(x) \geq 1$ that $y(x) \geq 1$ iff $x \geq 1$ or $x \leq -d/a$. Thus for positive x we have that $y(x) \geq 1$ if and only if $x \geq 1$. Writing $y(x)/x = \big(1+a(x-1)\big)/\big(1+(1-d)(x-1)\big)$ the proof follows simply, since $y(x)/x \geq 1$ if and only if $a(x-1) \geq (1-d)(x-1)$.

■

The condition that $a + d \leq 1$ can be seen as a burstiness condition. $a + d = 1$ gives Bernoulli arrivals, whereas for $a + d < 1$ the arrivals at successive times are positively correlated. This is just what is required for multiplexer modelling. The following theorem gives an exponential bound on the virtual waiting time for a given choice of x. In the next section we discuss the consequences of different choices of values for x.

Theorem 1. *Let $a + d \leq 1$ and $x \geq 1$. If $g(x) \geq 1$ then for any $b \geq 1$*

$$\mathbb{P}[Q \geq b] \leq x^{-1} f(x)^L y(x)^{1-b} \quad . \tag{2.3}$$

Proof: For $b \geq 1$

$$\{Q \geq b\} \subset \cup_{n \geq 1} \{Q_n \geq b, Q_{n-1} < b\} \subset \cup_{n \geq 1} B_n \cap B_{n-1}^c \quad ,$$

where $B_0 = \emptyset$ and $B_n = \{z_1 + \cdots + z_n - ns \geq b\}$ for $n \geq 1$. Now since $x \geq 1$, then for $n > 1$

$$B_n = \{x^{z_1 + \cdots + z_n} \geq x^{ns+b}\}$$

while

$$\begin{aligned} B_{n-1}^c &= \{z_1 + \cdots + z_{n-1} < (n-1)s + b\} \\ &= \{z_1 + \cdots + z_{n-1} \leq (n-1)s + b - 1\} \\ &\subseteq \{\big(y(x)/x\big)^{z_1 + \cdots + z_{n-1}} \geq \big(y(x)/x\big)^{(n-1)s+b-1}\} \end{aligned}$$

since by Proposition 2 $y(x)/x \leq 1$. Thus

$$\begin{aligned} B_n \cap B_{n-1}^c &\subset \{y(x)^{z_1 + \cdots + z_{n-1}} x^{z_n} \geq y(x)^{(n-1)s+b-1} x^{s+1}\} \\ &= \{Y_n \geq g(x)^{L(n-1)} x^{s+1} y(x)^{b-1}\} \\ &\subseteq \{Y_n \geq x^{s+1} y(x)^{b-1}\} \quad , \end{aligned}$$

the last inclusion since $g(x) \geq 1$. Furthermore, when $b \geq 1$

$$B_1 = \{z_1 \geq b + s\} = \{x^{z_1} \geq x^s x^b\} \subseteq \{x^{z_1} \geq x^{s+1} y(x)^{b-1}\} = \{Y_1 \geq x^{s+1} y(x)^{b-1}\}$$

Thus

$$\{Q \geq b\} \subset \cup_{n \geq 1} \{Y_n \geq y(x)^{b-1} x^{s+1}\} = \{\sup_{n \geq 1} Y_n \geq y(x)^{b-1} x^{s+1}\} \qquad (2.4)$$

Since Y is a positive martingale, the probability of the right hand side of (2.4) is bounded according to the maximal inequality for positive supermartingales (see [14]) by $y(x)^{1-b} x^{-(s+1)} \mathbb{E}[Y_1]$. Now $\mathbb{E}[Y_1] = \big((ax + d)/(a + d)\big)^L$, and so we obtain the upper bound (2.3).

■

3. Explicit bounds on tail probabilities.

We now turn to the question of finding the best bound out of those obtained for differing values of x in (2.3). In general this will depend of the value of b for which the bound is required. Looking at the extreme cases, then for $b = 1$ one wants to minimize the prefactor $x^{-1} f(x)^L$, whereas for $b \to \infty$ one wants to maximize the geometric decay constant $y(x)$. It turns out that by finding the x which minimizes $f(x)$ we obtain an explicit formula for the bound, whereas the maximum value for $y(x)$ is obtained only through a numerical search. In any case, first we must actually establish the existence of $x > 1$ such that $g(x) \geq 1$ as required in the proof of Theorem 1. This is done in the following proposition, the proof of which (being an intricate combination of simple convexity arguments) we defer to an appendix.

Theorem 2. *Let $a + d \leq 1$, $\sigma > a/(a + d)$ and $x > 1$.*

(1) If $f'(x) = 0$ then $g(x) > 1$.

(2) If $g(x) \geq 1$ then $f(x) \leq 1$.

The explicit bound is got as follows. Differentiating f then one sees that $f'(x) = 0$ when

$$x = x_\sigma := d\sigma/(a(1 - \sigma)) \quad .$$

By Thm. 2(1), $g(x_\sigma) > 1$, so that by Thm. 2(2), $f(x_\sigma) \leq 1$. So the prefactor in the bound (2.3) is less than 1. Inserting $x = x_\sigma$ in (2.3) gives in terms of

- the mean silence length $1/a$ units
- the mean burst length $1/d$ units
- the number of lines L
- the block work rate $s = \sigma L$ units

the explicit bound for the tail of the queue length distribution

$$\begin{aligned}\mathbb{P}[Q \geq b] &\leq \frac{1}{y(x_\sigma)^{b-1}x_\sigma}\left(\frac{ax_\sigma + d}{(a+d)(x_\sigma)^\sigma}\right)^L \\ &= \frac{a\Big(1-\big(a+\sigma(1-a-d)\big)\Big)}{d\big(a+\sigma(1-a-d)\big)}\left[\frac{(1-\sigma)\big(a+\sigma(1-a-d)\big)}{\sigma\Big(1-\big(a+\sigma(1-a-d)\big)\Big)}\right]^b \times \\ &\qquad \left[\frac{1}{a+d}\left(\frac{d}{1-\sigma}\right)^{1-\sigma}\left(\frac{a}{\sigma}\right)^\sigma\right]^L \end{aligned} \tag{3.1}$$

for any $b \geq 1$, provided that the burstiness condition $a + d \leq 1$ and the stability condition $\sigma > a/(a+d)$ are satisfied.

In fact $f(x_\sigma)^L$ is exactly the estimate on overflow probabilities found by Hui [8] for models of bufferless resources: in the present case this corresponds to bounding only $\mathbb{P}[Q \geq 1]$. This result finds the large deviation properties of the overflow probabilities in terms of the number of lines L. Our formula (3.1) can be seen as an extension of this to treat also the large deviation properties in the buffer size b, providing a simple interpolation between $b = 0$ and the asymptotic regime $b \to \infty$.

To find the largest possible decay constant requires a numerical search for the largest solution $x_{\max}$ of $g(x) = 1$, which gives $y(x_{\max})$ as the required value since y is increasing (as is verified by differentiation).

The use of the explicit bound (3.1) based on the choice $x = x_\sigma$ is clearly convenient, but it is worthwhile considering the circumstances under which it is accurate. In Table 1, for a multiplexer with service rate $s = 400$ of $L = 900$ sources each with mean activity $a/(a+d) = 0.4$, we have compared x_σ and $x_{\max}$ according to the burstiness of the source, as determined by $a+d$. Defining $b(x;\varepsilon)$ as the buffer size required for a loss probability ε using a bound based on x, we have compared $b(x_\sigma; 10^{-9})$ and $b(x_{\max}; 10^{-9})$. For highly bursty sources (i.e. $a + d$ very small) expected in ATM multiplexers, the two bounds are very close, the proportionate difference being most pronounced at $a+d = 1$ i.e. for Bernoulli arrivals. Less bursty

sources have been considered elsewhere [17]: use of a bound based on $x = x_{\max}$ seems to be more accurate in this case.

Whichever choice of x is made, one can straightforwardly obtain an upper bound on the mean queue length since

$$\mathbb{E}[Q] = \sum_{b \geq 1} b\mathbb{P}[Q = b] = \sum_{b \geq 1} \mathbb{P}[Q \geq b] \leq \frac{y(x)f(x)^L}{x(y(x) - 1)}$$

It is useful to know how this behaves as the system approaches overload, i.e. as σ approaches $a/(a + d)$. Taking $x = x_\sigma$ then one sees that x_σ and hence $y(x_\sigma)$ approach 1 in this limit, so that the bound diverges. Let us see more precisely in what manner this happens. From their definitions one sees that the factors $y(x_\sigma)/x$ and $f(x_\sigma)$ approach 1 from below, so we need consider only the asymptotics of $1/(y(x_\sigma) - 1)$. Setting $\varepsilon = \sigma - a/(a + d)$ we obtain

$$\mathbb{E}[Q] \leq \frac{1}{y(x_\sigma) - 1} = \frac{1}{\varepsilon} \frac{ad}{(a + d)^3} \left[1 + \varepsilon \frac{(a + d)^2}{d} + \varepsilon^2 \frac{(a + d)^2}{d^2}(1 - a - d)\right] \quad . \quad (3.2)$$

Note that (3.2) is exact: there are no higher order terms in ε.

One can compare this bound with Kingman's bound for independent arrivals [10] applied to a superposition of Bernoulli sources of the same mean activity $a/(a + d)$. In this case

$$\mathbb{E}[Q] \leq \frac{1}{\varepsilon} \frac{ad}{2(a + d)^2} \quad . \quad (3.3)$$

Comparing (3.2) and (3.3) we can say that the upper bound for the mean queue length in the Markovian case exceeds that for independent arrivals with the same mean activity by roughly a factor $2/(a + d)$, a quantity which is very large in the case that $a + d$ is very much less than 1.

4. Comparisons: the case $s = 1$; computer simulations.

In this section we compare our upper bounds with an explicit result for a special case of our model, and with computer simulations for more general cases.

The model for $s = 1$ has been treated by matrix methods in [17] (and developed in [13]) which make use of the particular form of the transition matrix on the combined state space of buffer and lines. An exact expression was obtained for the mean of the buffer delay, but not for its distribution. In terms of our parameters, the result of [17] for the mean queue length (rather than the mean delay) is $a^2(2 -$

$a-d)L(L-1)/(2(a+d)^2(a+d-aL))$, where $L < L_{max} = (a+d)/a$. Using $a = 0.095$ with $L_{max} = 10$, we found that the explicit bound over-estimated the exact expression by roughly a factor 2. The over-estimation was found to be smaller for models with larger values of L_{max}.

In computer simulations of larger systems we modelled lines with an activity of 0.4, taking $a = 0.03$ and $d = 0.045$. s was 40, giving $L_{max} = 100$. For the purposes of simulation these are scaled down by roughly an order of magnitude from typical projected values in ATM traffic.

Each simulation run comprised 10 million samples, each sample using $L < L_{max}$ calls to a Wichmann-Hill pseudo-random number generator[**18**]. Thus in each run there were not more than 10^9 calls in total, compared with a period of at least 10^{12} for the generator.

In Fig 1. and Fig. 2 the queue length distribution is given for loads 0.84 and 0.94 respectively. In both cases we see that the optimal bound leads to an over-estimation of the tail probabilities by much less than an order of magnitude. Note that the difference between the optimal bound and the explicit bound is small compared with the difference between the bounds and the simulation results. The optimal bound stays nearly parallel with the simulation curve for large buffer occupation, indicating that the decay constant is well estimated.

Appendix.

Proof of Theorem 2.

(1) Let $x = x_\sigma := d\sigma/(a(1-\sigma))$, which is seen by differentiation of f to be the solution of $f'(x) = 0$. Then $g(x) = x^\sigma/((1+a(x-1))^{1-\sigma}(1+(1-d)(x-1))^\sigma)$. By concavity of the function $r \mapsto h(r) := \log(1+r(x-1))$ then $g(x) \geq x^\sigma/(1+(a(1-\sigma)+\sigma(1-d))(x-1))$. A straightforward calculation shows that $a(1-\sigma)+\sigma(1-d) = \sigma((1-d)x+d)/x < \sigma$, so that

$$g(x) > \frac{(1+(x-1))^\sigma}{1+\sigma(x-1)} \quad .$$

With x now fixed, the right hand side of this equation is decreasing as a function of σ on $[0,1]$ and equal to 1 when $\sigma = 1$. Hence $g(x) > 1$.

(2) If $g(x) \geq 1$ then we find (by substitution using $y(x)$) that

$$\log f(x) \leq h(a/(a+d)) - (1-\sigma)h(a) - \sigma h(1-d) \quad . \qquad (A.1)$$

By the concavity of h, $h(a/(a+d))$ is smaller than $h(a)+(a/(a+d)-a)h'(a)$ and $h(1-d)+(a/(a+d)-(1-d))h'(1-d)$. Thus the right hand side of (A.1) is bounded above by

$$\frac{(1-\sigma)(a/(a+d)-a)}{1+a(x-1)}+\frac{\sigma(a/(a+d)-(1-d))}{1+(1-d)(x-1)} \quad .$$

By considering the numerator of this expression we see when $a+d\leq 1$ that this is non-positive if

$$v(x):=a/(a+d)-\sigma+(x-1)a(1-d-\sigma)/(a+d)\leq 0 \quad .$$

Now $g(x)\geq 1$ can be rewritten as $\sigma\geq h(a)/(h(1)+h(a)-h(1-d))$. Using the identity $\log B/\log A\geq (B-1)/(A-1)$ for $A>B>1$, and choosing $\log B=h(a)$ and $\log A=h(1)+h(a)-h(1-d)$, we find

$$\sigma-a/(a+d)\geq\frac{ad(x-1)(1-a-d)}{(ax+d)(a+d)} \quad . \qquad (A.2)$$

Upon substitution of (A.2) into the form of $v(x)$ we obtain precisely $v(x)\leq 0$, as required.

■

Acknowledgements.

N.G.D. thanks D.M. Corry and T. Curran of the Department of Electrical Engineering at Dublin City University for discussions on congestion in multiplexers, and M. Collier for his implementation of the pseudo-random number generator.

References.

[1] D. Anick, D. Mitra & M.M. Sondhi, Stochastic theory of a data-handling system with multiple sources, *Bell Sys. Tech. J.* **61**(1982) 1872-1894

[2] A. Baiocchi, N. Belafri Melazzi, M. Listanti, A. Roveri & R. Winkler, Loss performance analysis of an ATM multiplexer loaded with high speed on-off sources, *IEEE J. Select. Areas Commun.* **9**(1991) 388-393

[3] D.D. Botvich, Personal Communication,

[4] D.M. Corry, N.G. Duffield & T. Curran, Sub-burst mode queueing in an ATM voice multiplexer, Proceedings of 9th IEE Teletraffic Symposium Guildford, April 1992

[5] R.J. Gibbens & P.J. Hunt,, Effective Bandwidths for the multi-type UAS channel, *Queueing Systems* **9**(1991) 17-28

[6] J. Grandell, *Aspects of Risk Theory*, Springer Series in Statistics, Springer, New York (1991)

[7] H. Heffes & D.M. Lucantoni, A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance, *IEEE J. Select. Areas Commun.* 4(1991) 856-868

[8] J.Y. Hui, Resource allocation for broadband networks, *IEEE J. Selected Areas in Commun.* **6** (1988) 1598-1608

[9] J.F.C. Kingman, A martingale inequality in the theory of queues, *Proc. Camb. Phil. Soc.* **59**(1964) 359-361

[10] J.F.C. Kingman, Inequalities in the theory of queues, *J. Roy. Stat. Soc. Ser B* **32**(1970) 102-110

[11] L. Kosten, Stochastic Theory of data handling systems with groups of multiple sources, Proc. 2nd Int. Smyp. on the Performance of Computer Communication Systems, eds. H. Rudin & W. Bux, North-Holland, 1988

[12] R.M. Loynes, The stability of a queue with non-independent inter-arrival and service times, *Proc. Camb. Phil. Soc.* **58**(1962) 497-520

[13] M.F. Neuts, On Viterbi's formula for the mean delay in a queue of data packets, *Commun. Statist.-Stochastic Models* **6**(1990)87-98

[14] J. Neveu, *Discrete parameter martingales*, North-Holland, Amsterdam (1975).

[15] I. Norros, J.W. Roberts, A. Simonian & J.T. Virtamo, The superposition of variable bit rate sources in an ATM multiplexer , *IEEE J. Select. Areas Commun.* **9**(1991) 378-387

[16] K. Sriram & W. Whitt, Characterizing superposition arrival processes in packet multiplexers for voice data, *IEEE J. Select. Areas Commun.* 4(1986) 833-846

[17] A.M. Viterbi,, Approximate analysis of time synchronous packet networks, *IEEE J. Select. Areas Commun.* 4(1986) 879-890

[18] B.A. Wichmann & I.D. Hill, An efficient and portable pseudo-random number generator, *Appl. Stat.* **31**(1982) 188-190

Table 1
Comparison of bounds using x_σ and $x_{\max}$
for $L = 900$, $s = 400$, and activity $a/(a+d) = 0.4$

$a+d$	x_σ	$x_{\max}$	$b(x_\sigma; 10^{-9})$	$b(x_{\max}; 10^{-9})$
0.001	1.2	1.200	93763	93717
0.01	1.2	1.201	9377	9330
0.1	1.2	1.211	938	892
1	1.2	1.438	94	58

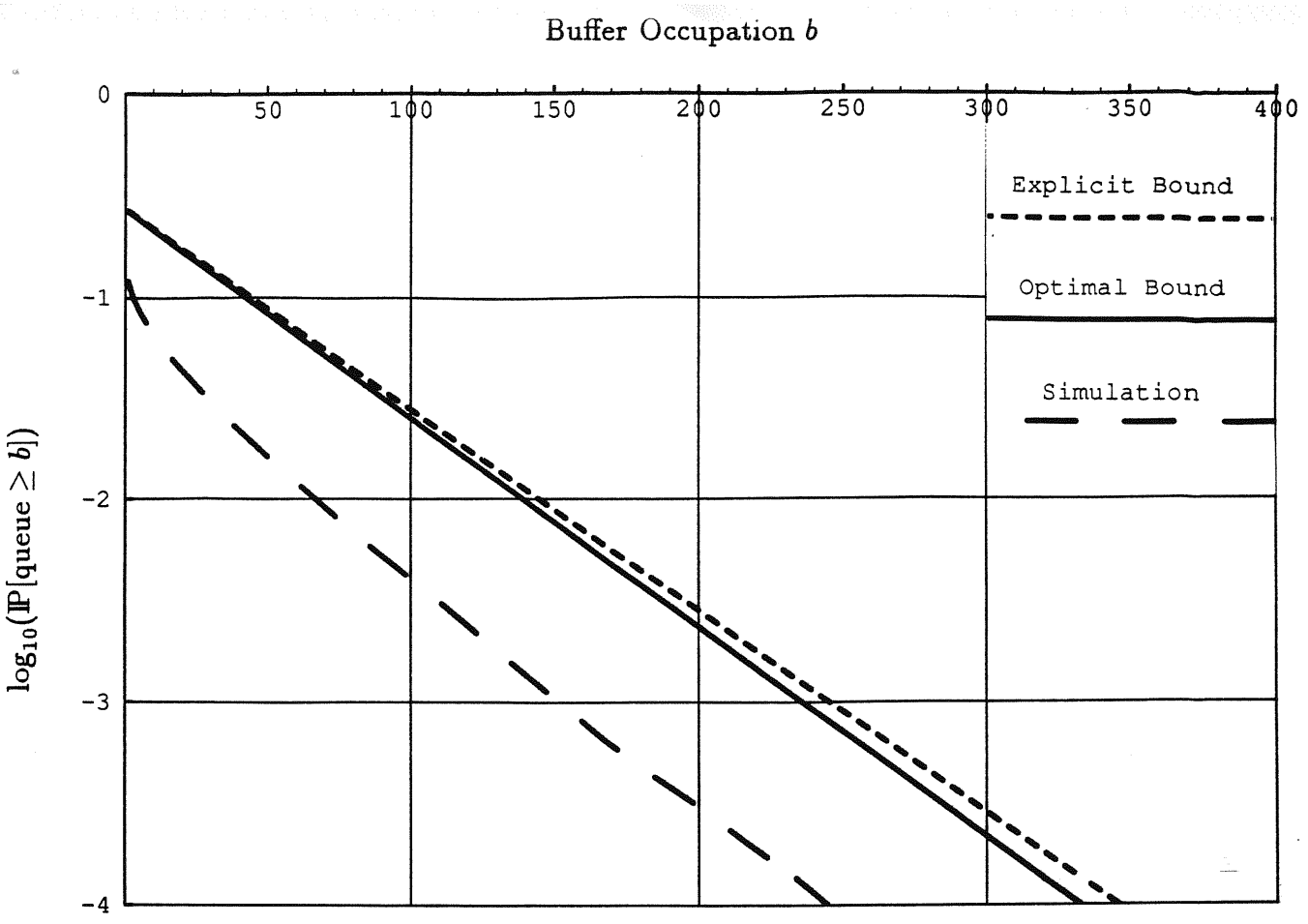

Fig. 1: Comparison of simulation and bounds for load $\rho = 0.84$

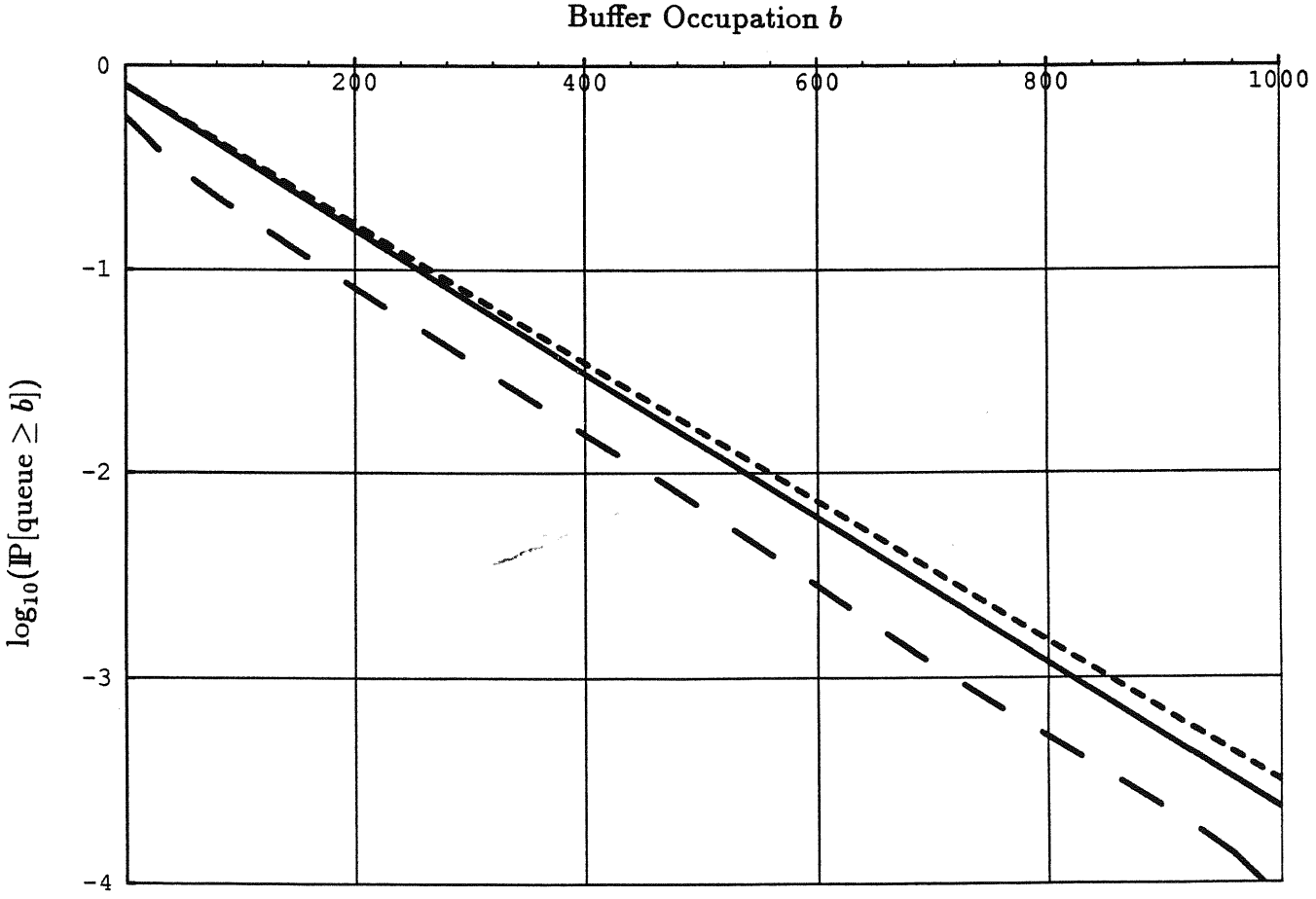

Fig. 2: Comparison of simulation and bounds for load $\rho = 0.94$